REMARKS

Claims 108-113 are pending after entry of the claim amendments requested herein. Claim 108 has been amended to correct obvious typographical/grammatical errors, broaden to recite the genus substituent "nucleobase," and in general comport more closely with the compounds which are the subject of the method claims presently pending before Examiner J. Riley, GAU 1637, in related application Ser. No. 10/072,975.

The outstanding issues in *this* case, responded to in order below, include objections predicated on alleged improper 1) shift in elected subject matter, 2) Abstract, 3) incorporation by reference and 4) rejections predicated on alleged indefiniteness of the terms "non-naturally occurring nucleobase (claim 108) and "phosphono peptide nucleic acid monomer" (claims 110 and 111).

1. Shift in elected subject matter

Claim 108 has been amended to recite the genus term "nucleobase," with provisional election still made with traverse to classic purines and pyrimidines species as noted in the specification and as otherwise understood in the art. As no prior art rejections have been advanced in this case, this genus "linking" claim is submitted to be appropriate. See also discussion points 3. and 4., *infra*.

2. Abstract

The Examiner requests the Abstract be amended to depict the compound of claim 108. Applicant accedes to the Examiner's request but asks that the change be made later, either 1) by Examiner's amendment upon his indication that claim 108 is allowable or 2) by Applicants upon notification of the same.

3. Incorporation by reference

The Examiner requests Applicants to amend the specification to include "essential" material allegedly omitted and only contained in documents incorporated by reference, e.g., definitions for "nucleobase" and "non-naturally occurring nucleobases." Applicants respectfully disagree for at least the following reasons.

The meanings of these terms are already sufficiently defined in the specification, e.g., on page 14, lines 3-26 and page 28 lines 3-26 (as set forth in Applicants' 26 July 2004 response) and

further on page 25, lines 20-26. In any event, these terms are also well-known in the art and as such need not be further defined. *See Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384, 231 U.S.P.Q. 81, 94 (Fed. Cir. 1986) ("[A] patent need not teach, and preferably omits, what is well known in the art"). The following illustrative exhibits attest to this:

Exh. A: Excerpts from Martin, D.W. et al., Harper's Review of Biochemistry, Twentieth Edition, Lange Publishers (1985). An index entry therein for "nucleobases" refers the reader to page 354 of a Chapter 25, entitled "Nucleotides," which section and chapter contains a hearty discussion of both natural and non-natural nucleobases;

Exh. B: World Wide Web printout definition for nucleobase: "a nitrogenous base that is a constituent of a nucleic acid, e.g., the purines: adenine, guanine, hypoxanthine, xanthine and the pyrimidines: cytosine, uracil, thymine."

Exh. C: Chaput J.C. and Switzer, C., "A DNA pentaplex incorporating nucleobase quintets," PNAS 96(19):10614-10619 (1999) (Abstract refers to the "nonstandard nucleobase" iso-guanine);

Exh. D: Bain, J.D. et al., Letters to Nature, 356: 537- 539 (1992) (Abstract refers to expansion of genetic code using "unnatural nucleoside bases");

Exh. E: Hoops, G.C. et al., "Template directed incorporation of nucleotide mixtures using azole-nucleobase analogs," Nucleic Acids Research, pp: 4866-4871 (1997) (see Abstract and Introduction reference to "artificial nucleobases").

To the extent the Examiner's opinion is predicated on genus:species considerations, the terms "nucleobases" and "non-naturally occurring nucleobases" are indeed genus and sub-genus terms, respectively, that subsume myriad individual species, many of which are already provided in the specification (see discussion, *supra*) and all of which need not be included expressly in the specification unless expressly claimed. *See Catalina Marketing Int'l, Inc. v. Coolsavings.Com, Inc.*, 289 F.3d 801 (Fed. Cir. 2002) (rejecting argument that use of "such as" in the specification with respect to illustrative species somehow renders a genus claim indefinite); *Utter v. Hiraga,* 845 F.2d 993, 998-99, 6 U.S.P.Q.2D 1709, 1714 (Fed. Cir. 1988) ("A specification may, within the meaning of § 112, contain a written description of a broadly claimed invention without describing all species that claim encompasses."); *In re Angstadt,* 537 F.2d 498, 502-03, 190 U.S.P.Q. 214, 218 (CCPA 1976) (Applicants "are not required to disclose every species encompassed by their claims even in an unpredictable art"); *In re Robins*, 57 C.C.P.A. 1321, 429

F.2d 452, 456-57, 166 U.S.P.Q. (BNA) 552, 555 (CCPA 1970) ("Mention of representative compounds encompassed by generic claim language clearly is not required by § 112 or any other provision of the statute. But, where no explicit description of a generic invention is to be found in the specification . . . mention of representative compounds may provide an implicit description upon which to base generic claim language.").

Viewed thusly, the terms in issue are not "essential" such that the application need be amended to incorporate further discussion. *See* MPEP § 608.01(p)(1)(a) (defining "essential" as "necessary to (1) describe the claimed invention, (2) provide an enabling disclosure of the claimed invention, or (3) describe the best mode..." If at a later date Applicants were to seek to introduce claims referencing species for which the application presently does not provide express support except by way of documents incorporated by reference, then that situation would admit of the propriety of such a requirement. However, that is respectfully submitted *not* to be the case here, at this time.

4. Definiteness

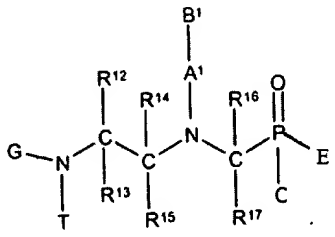
A. "non-naturally occurring nucleobase" in claim 113

For the reasons stated *supra*, the rejection with respect to claim 108 is now respectfully believed moot and with respect to claim 113 believed unsound and appropriately withdrawn.

B. "phosphono peptide nucleic acid monomer" in claims 110 and 111

The Examiner desires further clarification of what is meant by the term "phosphono peptide nucleic acid monomer." Applicants respond thusly:

In addition to the cyclic backbone species illustrated in structures II, III and IV on pages 20, 22 and 24 of the specification, it is clear that the term also embraces the linear structure



implicit in the dimers presented in structures VIII, IX, X and XI on pages 37, 40, 42 and 44 of the specification. Thus, Applicants respectfully submit that claims 110 and 111 *are* definite respecting the term "phosphono peptide nucleic acid monomer" and accordingly ask that the rejection be withdrawn.

5. Authorization to debit deposit account per 37 CFR 1.17(e)

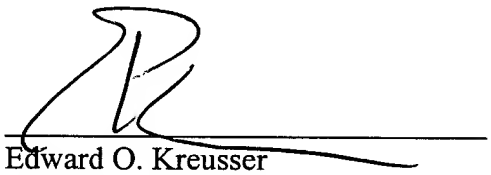
The undersigned hereby authorizes the PTO to debit deposit account # 502728 for the sum of $395 and to credit or debit said account for any other amounts that may be due or owing in connection with this matter.

CONCLUSION

Applicants respectfully submit that claims 108-113 are now in condition for allowance and they earnestly solicit a notice to that effect. Should any issues or questions remain, the Examiner is encouraged to telephone the undersigned at 858.485.0513 so that they may be promptly resolved without the need for an additional formal action and response thereto.

Dated: 11/10/04

Respectfully submitted,



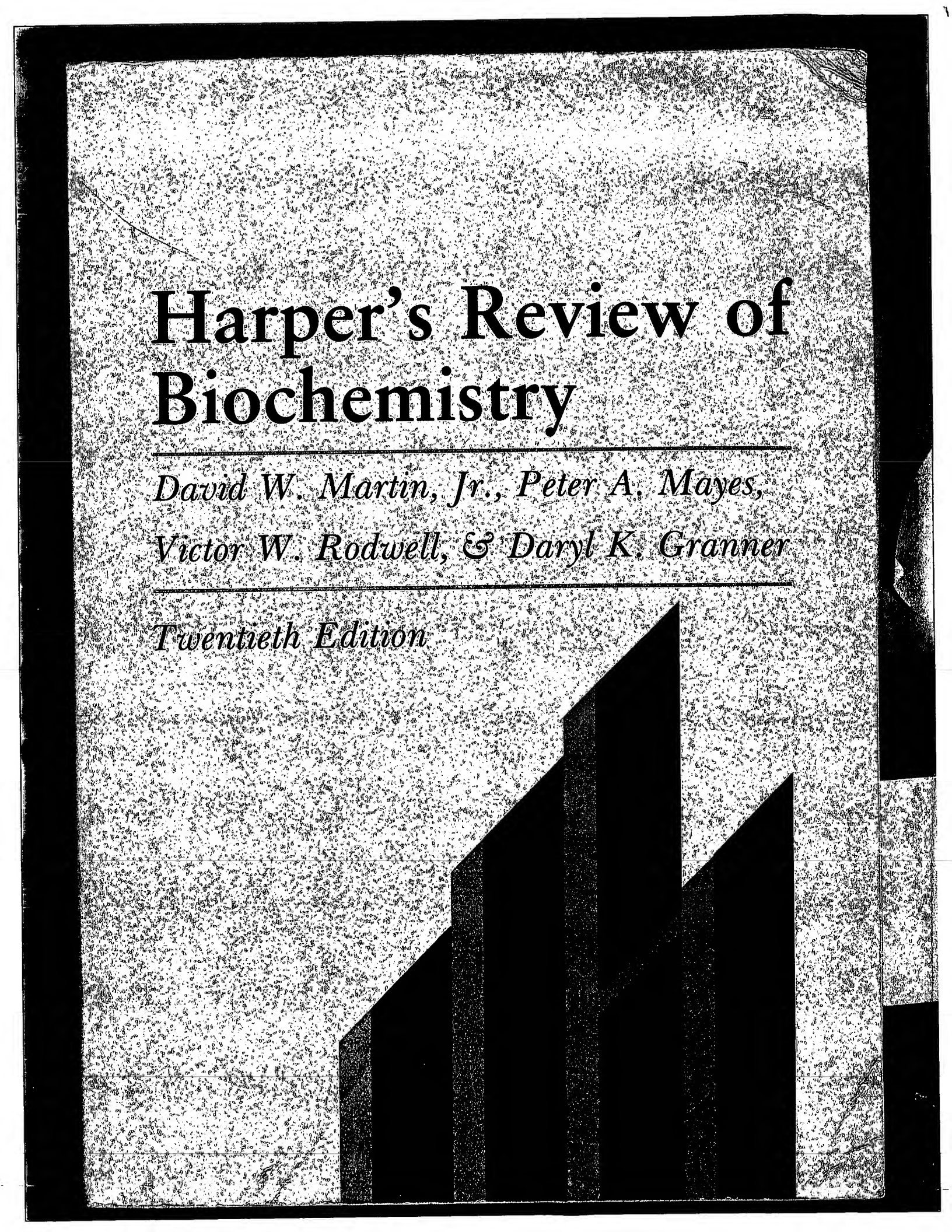
Edward O. Kreusser
Attorney for Applicants
BioTechnology Law Group
Reg. No. 38,523

Harper's Review of Biochemistry

David W. Martin, Jr., Peter A. Mayes,

Victor W. Rodwell, & Daryl K. Granner

Twentieth Edition



HARPER'S REVIEW OF BIOCHEMISTRY, 20th edition, is a sweeping revision of a popular work that has sold over a million copies worldwide since its first publication. Written for students in medicine, nursing, dentistry, veterinary science, and related fields, it assumes some minimal understanding of chemistry, including organic chemistry.

The twentieth edition continues to provide biologic, pathologic, and medical examples, the latter based on clinical experiences and abnormalities. Yet, keeping the student's needs in mind, the text is structured so that examples from medical situations may be easily skipped, if desired, without losing the opportunity to learn or solidify concepts. Over 700 illustrations.

Changes & Innovations in This Edition

- Hormone chapters totally rewritten with emphasis on molecular and biochemical aspects.
- New developments in regulation of gene expression.
- Up-to-date concepts in RNA processing to form mature mRNA.
- Recent developments in molecular basis of oncogenesis.
- Update on: cellular processes of protein secretion.
  fatty acid derivatives mediating inflammation.
  biochemical basis of blood clotting.

COMING IN 1985: BIOCHEMISTRY: A Synopsis, by Diane S. Colby

0-87041-038-5

Lange Medical Publications

Drawer L, Los Altos, California 94023

Italian Edition: *Piccin Nuova Libraria, S.p.A., Via Altinate, 107, 35121 Padua, Italy*
Spanish Edition: *Editorial El Manual Moderno, S.A. de C.V., Av. Sonora 206 - Col. Hipodromo, 06100-Mexico, D.F.*
French Edition: *Les Presses de l'Université Laval, Cité Universitaire, C.P. 2447, Quebec, Canada G1K 7R4*
Japanese Edition: *Maruzen Co., Ltd., P.O. Box 5050, Tokyo International 100-31, Japan*
Portuguese Edition: *Atheneu Editora São Paulo Ltda., Rua Marconi, 131 - 2.o andar, 01047 São.Paulo, Brazil*
Polish Edition: *Panstwowy Zaklad Wydawnictw Lekarskich, P.O. Box 379, 00-950 Warsaw 1, Poland*
German Edition: *Springer-Verlag GmbH & Co. KG, Postfach 10 52 80, 6900 Heidelberg 1, West Germany*
Indonesian Edition: *CV. E.G.C. Medical Publisher, P.O. Box 4276, 10711 Jakarta, Indonesia*
Serbo-Croatian Edition: *Savremena Administracija, Crnotravska 7-9, 11000 Belgrade, Yugoslavia*
Greek Edition: *Gregory Parisianos, 20, Navarinou Street, GR-106 80 Athens, Greece*

International Standard Book Number: *0-87041-038-5*
Library of Congress Catalogue Card Number: *61-10434*

Harper's Review of Biochemistry

Lithographed in USA

25 | Nucleotides

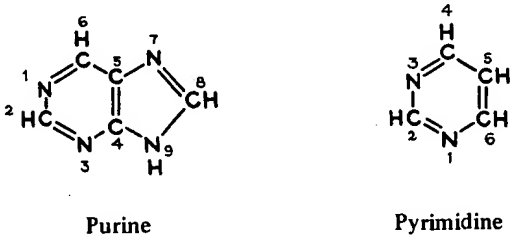
David W. Martin, Jr., MD

The nucleotides are important intracellular molecules of low molecular weight that participate in a wide variety of biochemical processes. Perhaps the best known role of the purine and pyrimidine nucleotides is to serve as the monomeric precursors of RNA and DNA. However, the **purine** ribonucleotides serve also in biologic systems as the ubiquitous high-energy source, ATP; as regulatory signals (cyclic AMP [cAMP] and cyclic GMP) in a wide variety of tissues and organisms; and as components of the widely used coenzymes FAD, NAD, and NADP and of an important methyl donor, S-adenosylmethionine.

The **pyrimidine** nucleotides, in addition to providing monomeric precursors for nucleic acids, serve as high-energy intermediates, such as UDP-glucose and UDP-galactose in carbohydrate metabolism and CDP-acylglycerol in lipid synthesis.

The various purine and pyrimidine bases that occur in the nucleotides are derived by appropriate substitution on the ring structures of the parent substances, purine or pyrimidine. Structures of these parent nitrogenous bases are shown in Fig 25–1. The positions on the rings are numbered according to the international system. Note that the direction of the numbering of the purine ring is different from that of the pyrimidine ring but that the number 5 carbon is the same in both heterocyclic compounds. Because of their π electron clouds, both the purine and pyrimidine bases are planar molecules, the significance of which is discussed in Chapter 27.

The 3 major **pyrimidine bases** found in the nucleotides of both prokaryotes and eukaryotes are **cytosine, thymine,** and **uracil** (Fig 25–2). The **purine bases, adenine** and **guanine,** are the 2 major purines found in living organisms. Two other purine bases, **hypoxanthine** and **xanthine,** also occur as intermediates in the metabolism of adenine and guanine (Fig 25–3). In humans, a completely oxidized purine base, **uric acid,** is formed as the end product of purine catabolism. This compound is discussed in greater detail in Chapter 26.



Purine Pyrimidine

Figure 25–1. Structures of purine and pyrimidine with the positions of the elements numbered according to the international system.

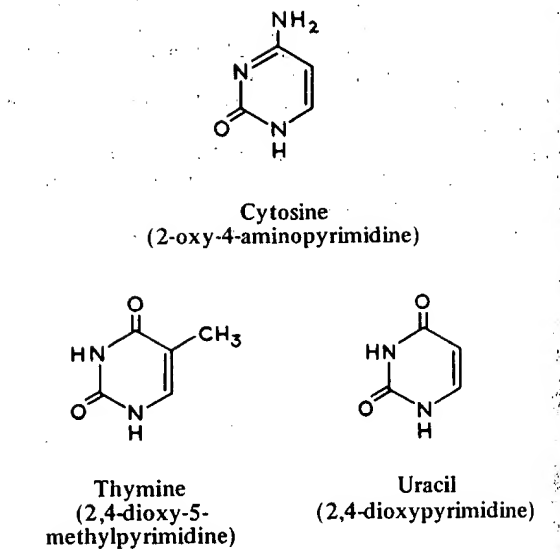
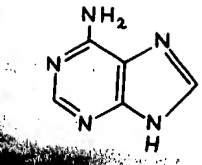


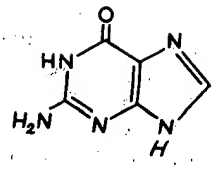
Figure 25–2. The 3 major pyrimidine bases found in nucleotides.

Because of keto-enol tautomerism, these aromatic molecules can exist in a lactim or lactam form (Fig 25–4); the latter is by far the predominant tautomer of guanine or thymine under physiologic conditions. (The importance of the lactim versus the lactam form becomes apparent in the discussions on base pairing and mutagenesis in Chapters 28 and 30.)

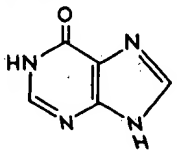
In plants, a series of purine bases containing methyl substituents occurs (Fig 25–5). Many have pharmacologic properties. Examples are coffee, which contains caffeine (1,3,7-trimethylxanthine); tea, which contains theophylline (1,3-dimethylxanthine); and cocoa, which contains theobromine (3,7-dimethylxanthine). The biologic properties of these



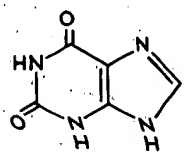
Adenine
(6-aminopurine)



Guanine
(2-amino-6-oxypurine)

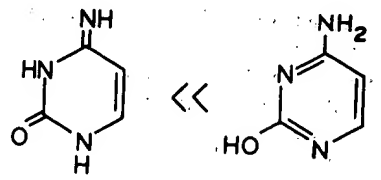


Hypoxanthine
(6-oxypurine)

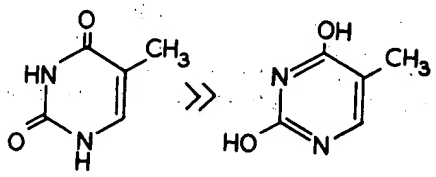


Xanthine
(2,6-dioxypurine)

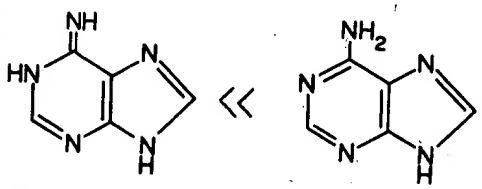
Figure 25–3. The major purine bases present in nucleotides.



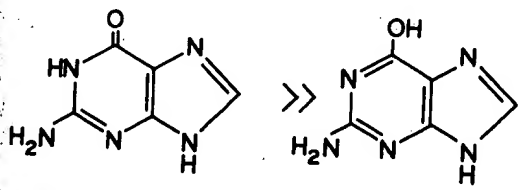
Cytosine (lactam) Cytosine (lactim)



Thymine (lactam) Thymine (lactim)

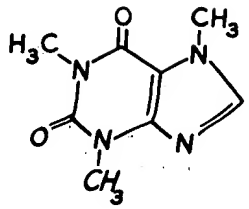


Adenine (lactam) Adenine (lactim)

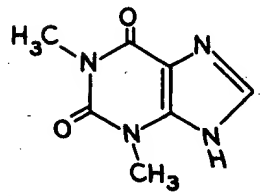


Guanine (lactam) Guanine (lactim)

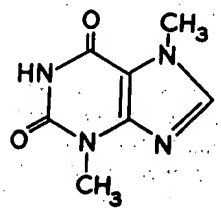
Figure 25–4. The structures of the tautomers of cystosine, thymine, adenine, and guanine with the predominant forms indicated.



Caffeine
(1,3,7-trimethylxanthine)



Theophylline
(1,3-dimethylxanthine)

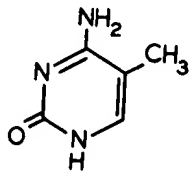


Theobromine
(3,7-dimethylxanthine)

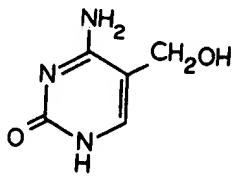
Figure 25–5. The structures of some methyl xanthines commonly occurring in foodstuffs.

compounds are described in Chapter 26 in the discussion of the metabolism of cyclic nucleotides.

In natural materials, numerous minor (ie, unusual) bases occur in addition to the 5 major bases—adenine, guanine, cytosine, thymine, and uracil—described above. Some of these unusual substituted bases are found only in the nucleic acids of bacteria and viruses, but many are also found in the DNA and transfer RNAs of both prokaryotes and eukaryotes. Both bacterial and human DNA contain, for example, significant quantities of 5-methylcytosine; bacteriophages contain 5-hydroxymethylcytosine (Fig 25–6). More recently, several unusual bases have been discovered in the messenger RNA molecules of mammalian cells. N^6-methyladenine, N^6,N^6-dimethyladenine, and N^7-methylguanine are found in the nucleic acids of mammalian cells (Fig 25–7). A uracil modified at the N_3 position by the attachment of an (α-amino,



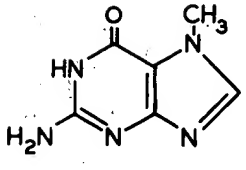
5-Methylcytosine



5-Hydroxymethylcytosine

Figure 25–6. The structures of 2 uncommon naturally occurring pyrimidine bases.

N^6,N^6-Dimethyladenine



N^7-Methylguanine

Figure 25–7. The structures of 2 uncommon naturally occurring purine bases.

α-carboxyl)-propyl group has also been detected in bacteria. The functions of these substituted purine and pyrimidine nucleotide bases are not fully understood.

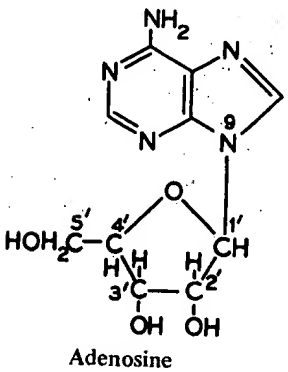
At neutral pH, guanine is the least soluble of the bases, followed in this respect by xanthine. Although uric acid as urate is relatively soluble at a neutral pH, its pK is 5.75, so that it becomes highly insoluble in a solution with a lower pH, such as urine. Guanine is not a normal constituent of human urine, but xanthine and uric acid do occur in human urine. In view of their low solubility, it is not surprising that these latter 2 purines are most likely to be found as constituents of stones formed within the urinary tract.

NUCLEOSIDES & NUCLEOTIDES

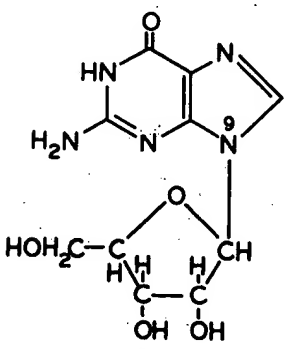
The **free** bases occurring in nature are much less abundant forms of purines and pyrimidines than are their nucleosides and nucleotides. A **nucleoside** (Fig 25–8) is composed of a purine or a pyrimidine base to which a sugar (usually either D-ribose or 2-deoxyribose) is attached in β-linkage at the N_9 or N_1, respectively. Thus, the adenine ribonucleoside **adenosine** consists of adenine with D-ribose attached at the 9 position. **Guanosine** consists of guanine with D-ribose attached at the 9 position. Cytidine is cytosine with ribose attached at its N_1 position. Uridine consists of ribose attached at the N_1 position of uracil.

The 2′-deoxyribonucleosides consist of 2-deoxyribose attached to the purine or pyrimidine bases at the same positions described above. The attachment of the ribose or 2-deoxyribose to the ring structures of the purine or pyrimidine bases is through an N-glycosidic bond, which is relatively acid-labile. Although theoretically, free rotation occurs about this N-glycosidic bond of the sugar moiety and the purine or pyrimidine ring structure, steric hindrance between these 2 moieties in fact hinders free rotation. In the naturally occurring nucleosides, the **anti** conformation is strongly favored over the **syn** form (Fig 25–9). As discussed in Chapter 27, the anti form is necessary for the proper positioning of the complementary purine and pyrimidine bases in the double-stranded B form of deoxyribonucleic acid. (Because of the conventional representation of the D-ribose, in most figures of this and other chapters, the purine and pyrimidine nucleosides and nucleotides are shown in the less favored syn conformation.)

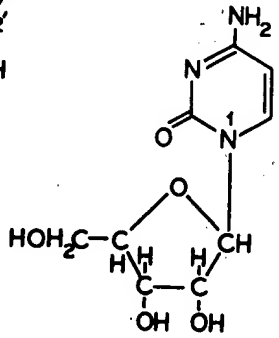
Nucleotides are nucleosides phosphorylated on one or more of the hydroxyl groups of the sugar (ribose or deoxyribose) (Fig 25–10). Thus, adenosine monophosphate (AMP or adenylate) is adenine + ribose + phosphate. 2′-Deoxyadenosine monophosphate (dAMP or deoxyadenylate) consists of adenine + 2-deoxyribose + phosphate. The only sugar com-



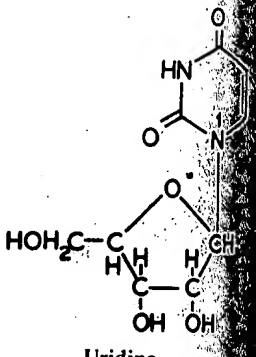
Adenosine



Guanosine



Cytidine



Uridine

Figure 25–8. Structures of ribonucleosides.

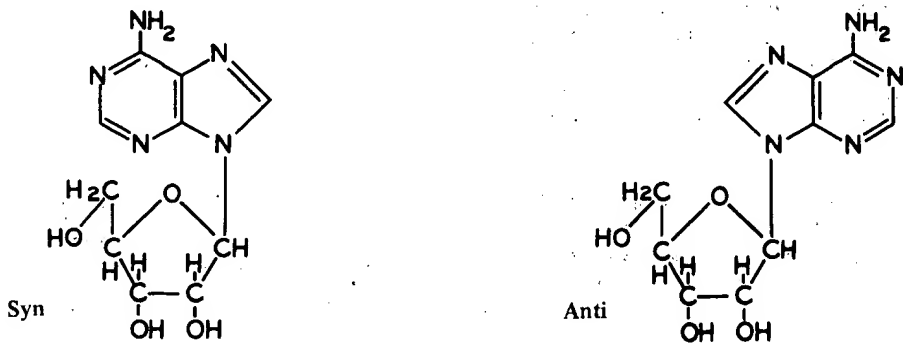


Figure 25–9. The structures of the syn and anti configurations of adenosine.

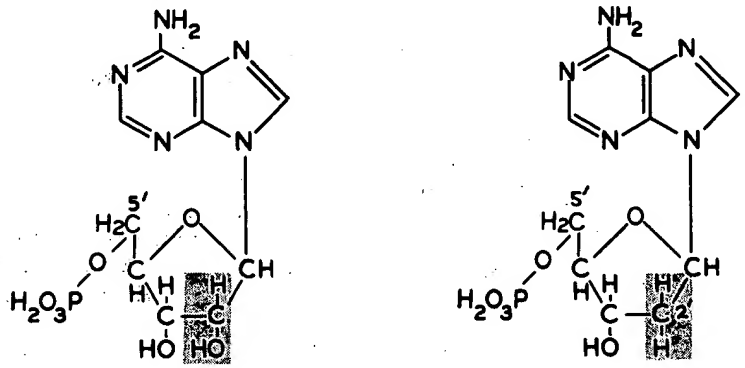


Figure 25–10. The structures of adenylic acid (AMP) *(left)* and 2'-deoxyadenylic acid (dAMP) *(right).*

monly found attached to uracil is ribose, and that commonly found attached to thymine is 2-deoxyribose. Therefore, thymidylic acid (TMP) is thymine + 2-deoxyribose + phosphate, and uridylic acid (UMP) is uracil + ribose + phosphate (Fig 25–11). DNA is a polymer of thymidylic acid, 2'-deoxycytidylic acid, 2'-deoxyadenylic acid, and 2'deoxyguanylic acid. RNA is a polymer containing uridylate, cytidylate, adenylate, and guanylate.

There are exceptions to the above structures of nucleotides. For example, in tRNA the ribose moiety is occasionally attached to uracil at the 5 position, thus establishing a carbon-to-carbon linkage instead of the usual nitrogen-to-carbon linkage. This unusual compound is called pseudouridine (Ψ). The tRNA molecules contain another unusual nucleotide structure, ie, thymine attached to ribose monophosphate. This compound is formed subsequent to the synthesis of the tRNA by methylation of the UMP residue by S-adenosylmethionine (see below). Pseudouridylic acid (ΨMP) is similarly rearranged from uridylic acid after the tRNA molecule has been synthesized.

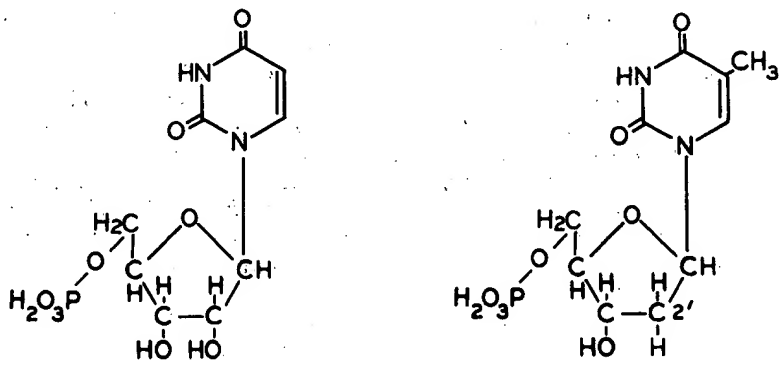


Figure 25–11. The structures of uridylic acid (UMP) *(left)* and thymidylic acid (TMP) *(right).*

Nomenclature of Nucleosides & Nucleotides

The position of the phosphate in the nucleotide is indicated by a numeral. For example, adenosine with the phosphate attached to carbon 3 of the sugar ribose would be designated adenosine 3′-phosphate. The prime mark after the numeral is required to differentiate the numbered position on the sugar moiety from the numbered position on a purine or pyrimidine base, which would not be followed by the prime mark. A nucleotide of 2′-deoxyadenosine with the phosphate moiety attached to the carbon 5 position of the sugar would be designated 2′-deoxyadenosine-5′-phosphate. (Fig 25–12.)

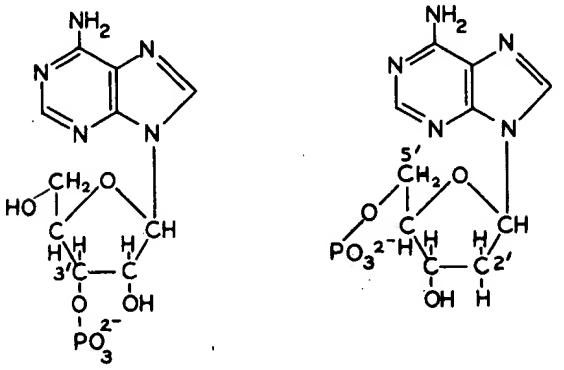


Figure 25–12. The structures of adenosine 3′-monophosphate *(left)* and 2′-deoxyadenosine-5′-monophosphate *(right)*.

The abbreviations A, G, C, T, and U may be used to designate a nucleoside in accordance with the purine or pyrimidine base it contains: adenine, guanine, cytosine, thymine, or uracil, respectively. The prefix d is added if the sugar of the nucleoside is 2′-deoxyribose. When the nucleoside occurs in the free form as a mononucleotide (ie, not a component of nucleic acid polynucleotide), the abbreviation MP (monophosphate) may be added to the abbreviation designating the nucleoside. For example, guanosine containing 2′-deoxyribose would be designated dG (deoxyguanosine) and the corresponding monophosphate with the phosphate esterified to the carbon 3 of the deoxyribose moiety is designated dG-3′-MP. Generally, when the phosphate is esterified to the carbon 5 of the ribose or deoxyribose moiety, the prefixed primed number (5′) is omitted. For example, guanosine 5′-monophosphate would be abbreviated GMP, while the 5′-monophosphate of 2′-deoxyguanosine would be designated dGMP. When 2 or 3 phosphates are attached to the sugar moiety in the acid anhydride form, the abbreviations DP (diphosphate) and TP (triphosphate) are added to the abbreviations for the corresponding purine or pyrimidine nucleoside. Thus, adenosine triphosphate with 3 phosphate residues attached to the 5′ carbon of the adenosine would be abbreviated ATP.

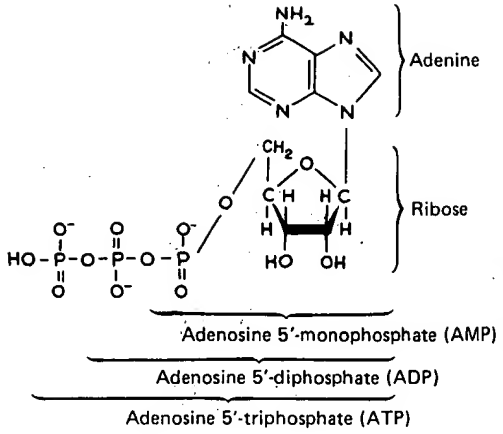


Figure 25–13. The structure of ATP and the structures of the corresponding diphosphate and monophosphate forms.

The structure of ATP is shown in Fig 25–13 along with its corresponding diphosphate and monophosphate forms. Because the phosphates are in the acid anhydride form—a low entropy situation—the phosphates are said to be high-energy ones, ie, high potential energy. The hydrolysis of 1 mol of ATP to ADP releases about 7 kcal of potential energy.

NATURALLY OCCURRING NUCLEOTIDES

Free nucleotides that are not an integral part of nucleic acids are also found in tissues. Many have important functions. Some of these compounds are briefly described.

Adenosine Derivatives

Adenosine diphosphate and adenosine triphosphate are important compounds in view of their participation in oxidative phosphorylation and, in the case of ATP, as the source of high-energy phosphate for nearly every energy-requiring reaction in the cell. The ATP concentration in most living mammalian cells is nearly 1 mmol/L. ATP is the most abundant intracellular free nucleotide.

Cyclic AMP (3′,5′-adenosine monophosphate; cAMP) is an important adenosine derivative that is present in most animal cells. cAMP mediates a series of diverse extracellular signals of considerable importance to the function of the organism as a whole. cAMP is formed from ATP (Fig 25–14). The reaction is catalyzed by the enzyme **adenylate cyclase,** the activity of which is regulated by a series of complex interactions many of which involve hormone receptors (see Chapter 35). cAMP is destroyed in tissues by its conversion to AMP in a reaction catalyzed by **cAMP phosphodiesterase.** Intracellular cAMP concentrations are usually near 1 μmol/L.

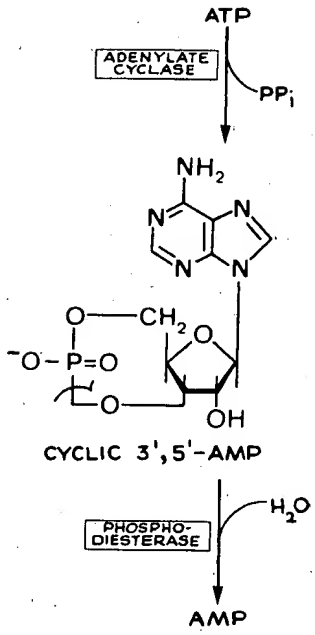


Figure 25–14. Formation of cAMP from ATP and destruction of cAMP by phosphodiesterase.

The incorporation of sulfate into ester linkages in compounds such as sulfated proteoglycans (see Chapter 32) requires the preliminary "activation" of the sulfate molecule. Sulfate is "activated" by reacting with ATP to form adenosine 3'-phosphate-5'-phosphosulfate (PAPS) in the reaction shown in Fig 25–15. The active sulfate moiety is also required as the substrate for sulfate conjugation reactions.

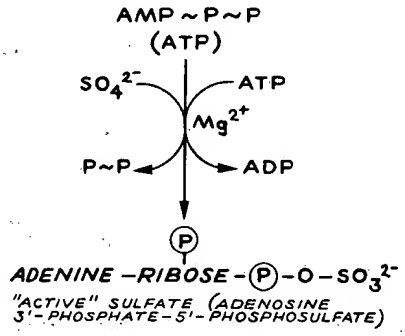


Figure 25–15. The formation of adenosine 3'-phosphate-5'-phosphosulfate.

Another important naturally occurring adenosine derivative, **S-adenosylmethionine** (Fig 25–16), serves as a form of "active" methionine. S-Adenosylmethionine serves widely as a methyl donor in many diverse methylation reactions and as a source of propylamine for the synthesis of polyamines.

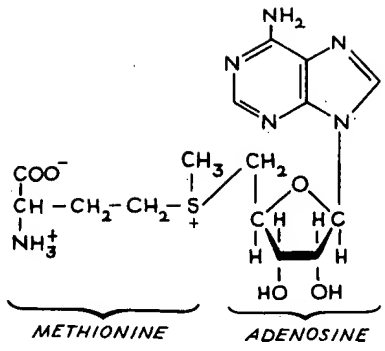


Figure 25–16. The structure of S-adenosylmethionine.

Guanosine Derivatives

Guanosine nucleotides, particularly guanosine diphosphate and guanosine triphosphate, serve in several energy-requiring systems. These are analogs of ADP and ATP, respectively. For example, the oxidation of α-ketoglutaric acid to succinyl-CoA in the tricarboxylic acid cycle involves oxidative phosphorylation with transfer of phosphate to GDP to form GTP. This phosphorylation reaction is quite similar to those involving the phosphorylation of ADP to ATP. GTP is required for the activation of adenylate cyclase by some hormones and serves both as an allosteric regulator and as an energy source for protein synthesis on polyribosomes. It therefore has an important role in the maintenance of the internal milieu.

Cyclic GMP (3',5'-guanosine monophosphate; cGMP [Fig 25–17]) appears also to be an important intracellular signal of extracellular events. In at least some cases, cGMP acts antagonistically to cAMP. cGMP is formed from GTP by an enzyme called **guanylate cyclase,** which is similar in many ways to adenylate cyclase. Guanylate cyclase, like adenylate cyclase, appears to be regulated by a variety of effectors, including hormones. cGMP is also catabolized by a phosphodiesterase to produce its respective 5'-monophosphate.

Hypoxanthine Derivatives

Hypoxanthine ribonucleotide, usually called inosinic acid (IMP, or inosinate in the salt form), is a

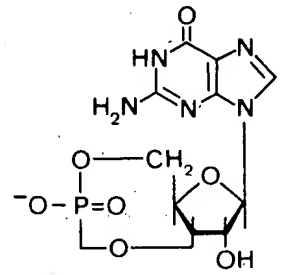


Figure 25–17. The structure of cyclic 3',5'-guanosine monophosphate (cyclic GMP; cGMP).

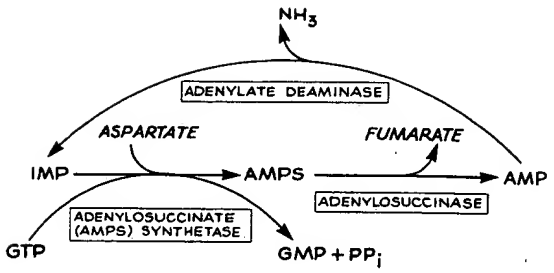


Figure 25–18. The purine nucleotide cycle.

precursor of all purine ribonucleotides synthesized de novo. Inosinate can also be formed by the deamination of AMP, a reaction which occurs particularly in muscle as part of the purine nucleotide cycle (Fig 25–18). Inosinate, derived from AMP, when reconverted to AMP results in the net production of ammonia from aspartate. Removal of the phosphate group from IMP forms the nucleoside inosine (hypoxanthine riboside), an intermediate in another cycle referred to as the purine salvage cycle (see Chapter 26).

Analogs of ADP and ATP in which the purine nucleoside derivative is inosine rather than adenosine have been found occasionally to participate in phosphorylation reactions. These compounds are inosine diphosphate (IDP) and inosine triphosphate (ITP).

Uracil Derivatives

Uridine nucleotide derivatives are important coenzymes in reactions involving the metabolism of hexoses and the polymerization of sugars to form starch and the oligosaccharide moieties of glycoproteins and proteoglycans. (See Chapter 33.) In these reactions, the substrates are uridine diphospho- sugars. For example, uridine diphosphoglucose (UDPGlc) is the precursor of glycogen. Another uridine nucleotide coenzyme, uridine diphosphoglucuronic acid (UDPGlcUA), serves as the "active" glucuronide for conjugation reactions such as the formation of bilirubin glucuronide (see Chapter 24).

Uracil also participates in the formation of high-energy phosphate compounds analogous to ATP, GTP, or ITP. Uridine triphosphate (UTP) is utilized, for example, in the reactions involving conversion of galactose to glucose in which the UDPGlc and UDP-Gal also are formed. UTP is the precursor for the polymerization of uridine nucleotides into RNA.

Cytosine Derivatives

Cytidine (cytosine-ribose) may form the high-energy phosphate compounds cytidine diphosphate (CDP) and cytidine triphosphate (CTP); the latter serves also as the precursor for the polymerization of CMP into nucleic acids. CTP is a nucleotide required for the biosynthesis of some phosphoglycerides in animal tissue. Reactions involving ceramide and CDP-choline are responsible for the formation of sphingomyelin and other substituted sphingosines.

Cyclic nucleotide derivatives of cytidine, analogous to those of adenosine and guanosine, have been described.

Vitamin Nucleotides

The functional moieties of many vitamins are nucleotides with structures analogous to purine and pyrimidine nucleotides. Riboflavin (vitamin B_2; see Chapter 10) functions as a ribitol 5′-phosphate derivative linked to AMP by a pyrophosphate bridge (FAD). Niacin is a constituent of 2 coenzymes, nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). In both of these cases, nicotinamide ribose phosphate is joined to an adenosine monophosphate through a pyrophosphate linkage. Coenzyme A is pantetheine linked to adenosine 3′-phosphate through a pyrophosphate moiety. One of the biologically active derivatives of cobalamin (vitamin B_{12}) requires the attachment of a 5′-deoxyadenosyl moiety through the 5′ carbon to the cobalt.

It should thus be clear that beyond providing the monomers of the structures of nucleic acids, the purine and pyrimidine nucleosides and nucleotides serve many diverse functions in living organisms.

SYNTHETIC DERIVATIVES

Synthetic analogs of nucleobases, nucleosides, and nucleotides are widely used in the medical sciences and clinical medicine. In the past, most of these uses have depended upon the role of nucleotides as components of nucleic acids for cellular growth and division. For a cell to divide, its nucleic acids must be replicated, requiring that the precursors of nucleic acids—the normal purine and pyrimidine deoxyribonucleotides—be readily available. One of the most important components of the oncologist's pharmacopeia is the group of synthetic analogs of purine and pyrimidine nucleobases and nucleosides.

The pharmacologic approach has been to use an analog in which either the heterocyclic ring structure or the sugar moiety has been altered in such a way as to induce toxic effects when the analog becomes incorporated into various cellular constituents. Many of these effects result from inhibition by the drug of specific enzyme activities necessary for nucleic acid synthesis or from the incorporation of metabolites of the drug into the nucleic acids where they alter the required base pairing essential to accurate transmission of information.

The most commonly used analogs of the purine or pyrimidine rings have substituents which do not occur naturally and which alter the base pairing or the interaction of the nucleotides with specific enzymes (Fig 25–19). Examples of these would be the 5-fluoro or 5-iodo derivatives of uracil or deoxyuridine, all of which serve as thymine or thymidine analogs, respectively. Both 6-thioguanine and 6-mercaptopurine, in which naturally occurring hydroxyl groups are replaced with thiol groups at the 6 position, are widely

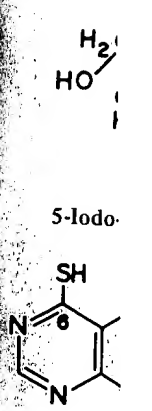


Figure 25–1
analogs (*abc*

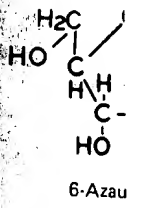
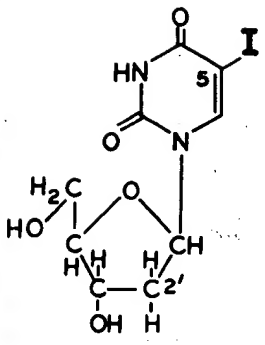
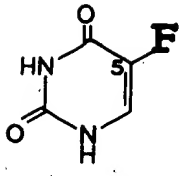


Figure 25–20.
8-azaguanine (

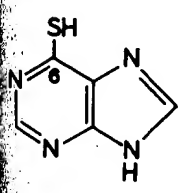
used clinicall
pyrimidine ri
5- or 6-azaur
(Fig 25–20).
The puri
(allopurinol) i
novo purine bi
used for the
Nucleosides c
the sugar mo
cytosine, Ara-
Ara-A), are u
viral infections
substances.)



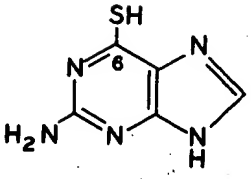
5-Iodo-2′-deoxyuridine



5-Fluorouracil

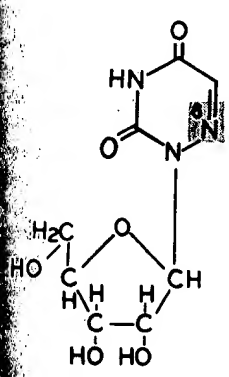


6-Mercaptopurine

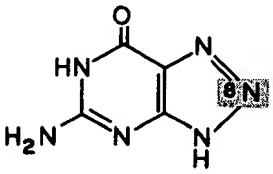


6-Thioguanine

Figure 25–19. The structures of 2 synthetic pyrimidine analogs *(above)* and 2 synthetic purine analogs *(below).*



6-Azauridine

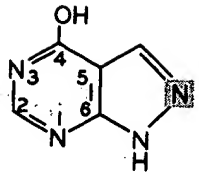


8-Azaguanine

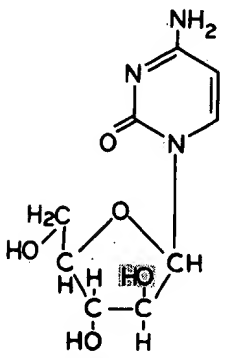
Figure 25–20. The structures of 6-azauridine *(left)* and 8-azaguanine *(right).*

used clinically. The analogs in which the purine or pyrimidine ring contains extra nitrogen atoms, such as 5- or 6-azauridine or azacytidine and 8-azaguanine (Fig 25–20), also have been tested clinically.

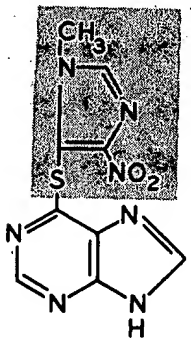
The purine analog 4-hydroxypyrazolopyrimidine (allopurinol) is widely marketed as an inhibitor of de novo purine biosynthesis and of xanthine oxidase. It is used for the treatment of hyperuricemia and gout. Nucleosides containing arabinose rather than ribose as the sugar moieties, notably cytarabine (arabinosyl cytosine, Ara-C) and vidarabine (arabinosyl adenine, Ara-A), are used in the chemotherapy of cancer and viral infections. (See Fig 25–21 for structures of these substances.)



Allopurinol (lactim).



Arabinosyl cytosine



Azathioprine

Figure 25–21. The structures of 4-hydroxypyrazolopyrimidine (allopurinol), arabinosyl cytosine (cytarabine), and azathioprine.

Azathioprine, which is catabolized to 6-mercaptopurine, is useful in organ transplantation as a suppressor of events involved in immunologic rejection. A series of nucleoside analogs with antiviral activities has been studied for several years; one, 5-iododeoxyuridine (see above), has been demonstrated to be effective in the local treatment of herpetic keratitis, an infection of the cornea by herpesvirus.

Numerous analogs of purine and pyrimidine ribonucleotides have been synthesized so as to generate nonhydrolyzable di- or triphosphates for use in vitro. These analogs allow the investigator to determine whether given biochemical effects of nucleoside di- or triphosphates require hydrolysis or whether their effects are mediated by occupying specific nucleotide binding sites on enzymes or regulatory proteins. Fig 25–22 depicts 2 such analogs of guanosine triphosphate.

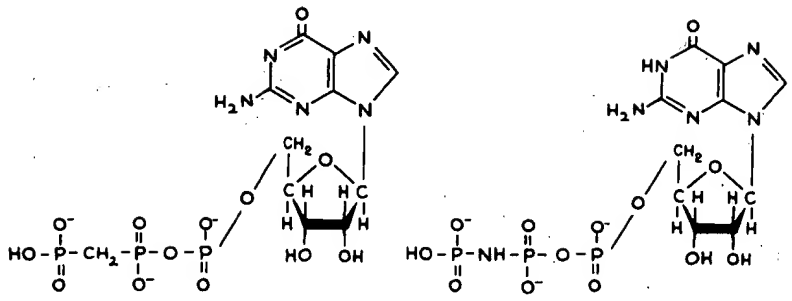


Figure 25–22. β-γ Methylene and β-γ imino derivatives of GTP, which cannot be hydrolyzed between the β and γ phosphates.

• • •

References

Henderson JF, Paterson ARP: *Nucleotide Metabolism: An Introduction*. Academic Press, 1973.

Michelson AM: *The Chemistry of Nucleosides and Nucleotides*. Academic Press, 1963.

Prusoff WH, Ward DC: Nucleoside analogs with antiviral activity. *Biochem Pharmacol* 1976;**25**:1233.

Mouse Genome Informatics
MGI Home Help

Search for
Go
in these sections
All sections
Gene symbols/names
Accession IDs
Phenotypes
Gene Expression

Advanced search for...

Search Categories

All Search Tools
Genes/Markers
Alleles/Phenotypes
Strains/Polymorphisms
Expression
Sequences
Comparative Maps/Data
Mouse Maps/Data
Mouse Tumor Biology
Probes/Clones
References
Vocabulary Browsers
Gene Ontology (GO)
Anatomical Dictionary
Phenotype Classifications

MouseBLAST

IMSR (Find Mice) NEW!

Additional Resources

Citing These Resources
Funding Information
Warranty Disclaimer
& Copyright Notice
Send questions and comments to User Support.

The Jackson Laboratory
last database update

Gene Ontology Browser

Term Detail

GO term: **nucleobase metabolism**
GO id: **GO:0009112**
Definition: **The chemical reactions and physical changes involving a nucleobase, a nitrogenous base that is a constituent of a nucleic acid, e.g. the purines: adenine, guanine, hypoxanthine, xanthine and the pyrimidines: cytosine, uracil, thymine.**
Number of paths to term: **3**

Ⓘdenotes an 'is-a' relationship
Ⓟdenotes a 'part-of' relationship

Gene_Ontology
Ⓟbiological_process
Ⓘphysiological process
Ⓘmetabolism
Ⓘaromatic compound metabolism
Ⓘaromatic amino acid family metabolism +
Ⓘaromatic compound biosynthesis +
Ⓘaromatic compound catabolism +
Ⓘbenzene and derivative metabolism +
Ⓘbenzyl isoquinoline alkaloid metabolism +
Ⓘcoumarin metabolism +
Ⓘcytokinin metabolism +
Ⓘectoine metabolism +
Ⓘenterobactin metabolism +
Ⓘfolic acid and derivative metabolism +
Ⓘindole and derivative metabolism +
Ⓘmandelate metabolism +
Ⓘnucleobase metabolism [GO:0009112] *(13 genes, 14 annotations)*
Ⓘnucleobase biosynthesis +
Ⓘnucleobase catabolism +
Ⓘpurine base metabolism +
Ⓘpyrimidine base metabolism +
Ⓘommochrome metabolism +
Ⓘphenol metabolism +
Ⓘphenylpropanoid metabolism +
Ⓘpteridine and derivative metabolism +
Ⓘpyridine metabolism +
ⒾS-adenosylmethioninamine metabolism +
ⒾS-adenosylmethionine metabolism +
Ⓘs-triazine compound metabolism +
Ⓘsalicylic acid metabolism +

PNAS
Proceedings of the National Academy of Sciences of the United States of America

Proc Natl Acad Sci U S A. 1999 September 14; 96 (19): 10614–10619
Biochemistry

A DNA pentaplex incorporating nucleobase quintets

John C. Chaput and Christopher Switzer†

Department of Chemistry, University of California, Riverside, CA 92521

Edited by Leslie Orgel, The Salk Institute for Biological Studies, San Diego, CA, and approved July 7, 1999, (received for review April 29, 1999)

† To whom reprint requests should be addressed. E-mail: switzer@citrus.ucr.edu.

Abstract

Supramolecular self-assembly is an integral step in the formation of many biological structures. Here we report a DNA pentaplex that derives from a metal-assisted, hydrogen bond-mediated self-assembly process. In particular, cesium ions are found to induce pentameric assembly of DNA bearing the nonstandard nucleobase iso-guanine. The pentaplex was designed by using a simple algorithm to predict nucleobase structural requirements within a quintet motif. The design principles are general and should extend to complexes beyond pentaplex. Structures exhibiting molecularities of five or more were previously accessible to peptides, but not nucleic acids.

Introduction

In abiologic systems, metal-mediated self-assembly has yielded arrays (1), nanometer-sized dendrimers (2), and other systems (3), notably, double- and triple-helical metal complexes that bear a formal resemblance to DNA (4). Recognition in the latter is driven by ligand constraints and metal coordination geometry. Recognition within DNA double and triple helices, in contrast, depends primarily on hydrogen bond complementarity. Quadruple helices of DNA based on the G-quartet motif are unusual, however, in that they rely on both hydrogen bonding and metal ion coordination (5, 6). Here we demonstrate the ability to expand DNA molecularity beyond quadruplexes by engineering nucleobases to fit dimensions required of higher-order motifs. Specifically, we design and characterize a DNA pentaplex.

nature

nature 09 April 1992

letters to nature

Nature **356**, 537 - 539 (1992); doi:10.1038/356537a0

Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code

J. D. Bain, Christopher Switzer, Richard Chamberlin & Steven A. Bennert

ONE serious limitation facing protein engineers is the availability of only 20 'proteinogenic' amino acids encoded by natural messenger RNA. The lack of structural diversity among these amino acids restricts the mechanistic and structural issues that can be addressed by site-directed mutagenesis. Here we describe a new technology for incorporating non-standard amino acids into polypeptides by ribosome-based translation. In this technology, the genetic code is expanded through the creation of a 65th codon-anticodon pair from unnatural nucleoside bases having non-standard hydrogen-bonding patterns[1,2]. This new codon-anticodon pair efficiently supports translation *in vitro* to yield peptides containing a non-standard amino acid. The versatility of the ribo-some as a synthetic tool offers new possibilities for protein engineering, and compares favourably with another recently described approach in which the genetic code is simply rearranged to recruit stop codons to play a coding role[3–9].

Nucleic Acids Research

Nucleic Acids Research | Pages 4866-4871

Template directed incorporation of nucleotide mixtures using azole-nucleobase analogs

Geoffrey C. Hoops, Peiming Zhang[1], W. Travis Johnson, Natasha Paul, Donald E. Bergstrom[1] and V. Jo Davisson*

Department of Medicinal Chemistry and Molecular Pharmacology, 1333 Robert Heine Pharmacy Building, Purdue University, West Lafayette, IN 47907-1333, USA and [1]Walther Cancer Institute, Indianapolis, IN 46208, USA

Received November 4, 1997; Accepted November 5, 1997

ABSTRACT

DNA that encodes elements for degenerate replication events by use of artificial nucleobases offers a versatile approach to manipulating sequences for applications in biotechnology. We have designed a family of artificial nucleobases that are capable of assuming multiple hydrogen bonding orientations through internal bond rotations to provide a means for degenerate molecular recognition. Incorporation of these analogs into a single position of a PCR primer allowed for analysis of their template effects on DNA amplification catalyzed by *Thermus aquaticus* (*Taq*) DNA polymerase. All of the nucleobase surrogates have similar shapes but differ by structural alterations that influence their electronic character. These subtle distinctions were able to influence the *Taq* DNA polymerase dependent incorporation of the four natural deoxyribonucleotides and thus, significantly expand the molecular design possibilities for biochemically functional nucleic acid analogs.

INTRODUCTION

Modified nucleobases that can function by degenerate recognition of natural nucleic acids would be invaluable tools for nucleic acid manipulation and applications in protein engineering. Nucleobases that show loss of discrimination when participating in DNA replication are rare and include only a few purine or pyrimidine derivatives. 8-Hydroxyguanine (1), 2-hydroxyadenine (2), 6-*O*-methylguanine (3-5) and xanthine (4,6) direct the incorporation of (C and A), (T and A), (T and C) and (T and C), respectively, in ratios that are highly polymerase dependent. Pyrimidine analogs O^2-ethylthymidine and O^4-ethylthymidine have been shown to direct the incorporation of A and T or A and G using the Klenow DNA polymerase I (7). Rationally designed non-discriminate bases such as 2-amino-6-methoxyaminopurine (K base), 6*H*, 8*H*-3,4-dihydropyrimido[4,5-*c*][1,2]oxazin-7-one (P base) (8), and N^6-methoxyadenine (Z base) (9,10) have provided oligonucleotide sequences with relaxed base pairing specificity when assessed in melting studies. However, when these tautomeric bases are used in DNA templates for enzyme catalyzed replication, they preferentially direct the incorporation of just one natural purine or pyrimidine deoxyribonucleotide (11).

Our approach toward the development of degenerate nucleic acid bases is founded on the principles of steric and conformational freedom for critical molecular recognition elements. Illustrated in Figure 1 is the conceptual evolution of a series of 1,3-azole-carboxamide heterocycles from inosine, a nucleotide that can base pair with A and C. A family of potential nucleobase analogs are represented within the general substituted azole heterocycles; as shown in Figure 1 this subclass includes eight different structures. A range of hydrogen bonding donor and acceptor patterns mediating nucleobase pairing are possible for these azoles as displayed in Figure 2. Internal bond rotations about the glycosidic bond and carboxamide side chain define alternative placement of ring nitrogens through which the azolecarboxamides can mimic the hydrogen bonding patterns of each of the natural nucleobases.

Figure 1. (**A**) Origin of the general structural subclass of azole nucleobases. (**B**) Structures of azole deoxyribonucleosides used in this study.

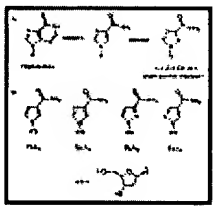
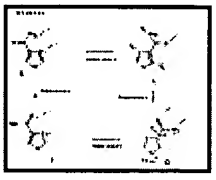


Figure 2.Hypothetical recognition features of the azole nucleobase analogs. Through alternative placements of ring nitrogens and conformational changes about bonds a and b, azolecarboxamides can mimic the in-plane hydrogen bonding patterns in nucleic acids with each of the natural nucleobases. A varied range of electronic configurations for DNA polymerase interactions are also possible at the arrow ([dArr]) indicated positions.



Experimental evidence that a theoretical hydrogen bonding analysis may be predictive has been obtained for imidazole-4- carboxamide **ImA$_4$** (Fig. 1 B). Base pairs in which **ImA$_4$** pairs with T or G are structurally most plausible. The results of T_m studies on oligonucleotide sequences containing **ImA$_4$** opposite T or opposite G show that these are significantly more stable than the corresponding sequences in which this base is placed opposite A or C (12). While these base pairing properties of the azole bases are of use in describing nucleic acid hybridization and structure, they are not complete when considering

their potential in enzyme catalyzed events such as DNA replication. Several critical features for enzyme recognition can be conceived within these simple azole bases as postulated in Figure 2. The potential enzyme recognition features of these heterocycles are indicated as the endocyclic nitrogens and the carboxamide side chain which also have alternative placements in the active site through simple bond rotations.

The studies presented herein focus upon a series of substituted 5-member azole heterocycles (as defined by the general structure in Fig. 1 A) that serve as nucleobase analogs in oligonucleotide templates for DNA amplification (Fig. 1 B). This study is the first report on the ability of the carboxamide group on simple azole rings to direct the incorporation of deoxyribonucleotides (dNTP) into replicating DNA strands catalyzed by a DNA polymerase. *Taq* DNA polymerase (13) was selected for these studies since it lacks a 3'-5' proof reading exonuclease activity (14) and permits an assessment of the directing properties of four nucleobase analogs. In addition, an assay system has been developed to evaluate the potential of PCR primers containing representative azole carboxamide bases to create sequence degeneracy under suitable conditions for *in vitro* amplification of DNA using a thermostable polymerase.

MATERIALS AND METHODS

Preparation of nucleosides and oligonucleotides

A full account of the syntheses of the nucleobase analogs phosphoramidite precursors for the oligonucleotides in Table 1 will be the subject of forthcoming articles. The 2'-deoxyribonucleosides containing **PrN_3**, **PrA_3**, **ImA_4** and **PzA_3** were synthesized as previously described (15-17). Addition of the 5'-dimethoxytrityl protecting group followed by the 3'-phosphoramidite to the 2'-deoxyribonucleosides provided precursors for oligonucleotide synthesis. The A, C, G, T and I phosphoramidites were purchased from Biogenics. The abasic phosphoramidite was obtained from Glenn Research. A preparation of the nucleoside corresponding to **PzA_4** followed from the sodium salt of ethyl pyrazole-4-carboxylate which was glycosylated with [alpha]-1-chloro-O^3,O^5-ditoluoyl-2'-deoxyribose. Treatment of the glycosylation product with methanolic ammonia at 120°C for 2 weeks afforded 1-([beta]-d-2'-deoxyribosyl)-pyrazole-4-carboxamide, **PzA_4**. The anomeric configuration of **PzA_4** was assigned as [beta] on the analysis of the ^{1}H-NMR signals for the anomeric and 2' protons of the corresponding 5'-dimethoxytrityl derivative (18,19) [^{1}H-NMR (p.p.m. in DMSO-d^6): 2.22-2.27 (1H, m, H-2''), 2.58-2.62 (1H, m, H-2'), 6.13 (1H, dd, $J_{1',2'}$ = 6.5 Hz, $J'_{14,24}$ = 4.5 Hz, H-1')].

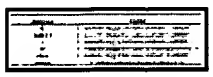


Table 1. Oligonucleotide sense strand primers for PCR of *E.coli hisF* G, guanine; #, Azole 1-5which include the deoxyribonucleotides of **PrN_3**, **PrA_3**, **PzA_4**, **ImA_4**, **PzA_3**, respectively (Fig. 1); I, hypoxanthine; M, a mixture of the four naturally occurring bases A, C, G and T; B, an abasic site.

Sense strand 40mer primers for PCR corresponding to positions -10 to 30 of the *Escherichia coli hisF* gene in the expression vector p*hisF-tac* (Table 1) were prepared. The 30mer antisense primer was composed entirely of natural bases (Table 1). All oligonucleotide primers were prepared either (i) using an ABI 380B or 392 DNA synthesizer in the Laboratory for Macromolecular Structure at Purdue University, or (ii) by Midland Certified Reagent Company (Midland, TX) using phosphoramidites

described above. The PCR primer containing an equal distribution of purine/pyrimidine bases at position #36 was prepared by mixing an equal amount of four primers. All primers were purified by preparative gel electrophoresis under denaturing conditions (20) (7 M urea, 50-55°C) on 20% (19% acrylamide, 1% *N,N'*-methylenebisacrylamide) polyacrylamide gels (1.4 mm * 30 cm * 40 cm) in TBE (90 mM Tris-borate, 2 mM EDTA) buffer on a Gibco BRL model S2 apparatus at 1000 V. Crude oligonucleotide (trityl off) from a 200 nmol scale synthesis was dissolved in 50 µl TE8 (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and diluted 2-fold by addition of 50 µl 2* formamide load buffer (80% formamide, 10 mM EDTA, 0.05% bromophenol blue). Gels were preequilibrated at 1000 V for 1-2 h before loading sample. After sample loading, gels were run at constant voltage until the bromophenol blue tracking dye had moved through 50-75% of the gel length. The gel was removed from the glass sandwich and bands were visualized by shadowing with a handheld short-wavelength UV (ultraviolet) lamp. The target oligonucleotide band was excised with a sharp scalpel. The purified oligonucleotide was isolated from the polyacrylamide gel using the `crush and soak' method (20,21), followed by phenol/chloroform extraction and ethanol precipitation. The purified primer was resuspended in TE8 (150 µl). Primer concentrations were estimated from UV absorbance at 260 nm as 200-fold dilutions in TE8 on a Cary 3 UV-visible spectrophotometer (Varian).

PCR

The 40mer oligonucleotides containing modified bases (Table 1) served as sense strand primers for polymerase chain reactions (PCRs) catalyzed by the *Taq* DNA polymerase (Amplitaq from Perkin-Elmer). Template DNA for PCR was produced by *Pvu*II digestion of the expression vector p*hisF-tac* (22). The following conditions for PCR were found to consistently and accurately amplify the template DNA containing the *hisF* gene: 100 µl reactions containing 100 mM Tris (pH 7.8 at 25°C), 50 mM KCl, 10 µg/ml gelatin, 1.5 mM $MgCl_2$, 100 pmol sense-strand primer, 100 pmol antisense-strand primer, fmol quantities of template DNA and 2.5 U Amplitaq^{tm} in 500 µl microcentrifuge tubes are topped with 75 µl mineral oil and subjected to 31 thermal cycles of PCR (95°C, 1 min; 37°C, 1 min; 70°C, 5 min; repeat) followed by an additional 10 min extension at 70°C in a PTC-100 thermocycler (MJ Research). Prior to purification, an initial analysis for successful amplification of the template DNA was made by agarose gel electrophoresis (21) of a portion (5 µl) of the dsDNA product mixture against molecular weight standards (Sigma). The products were purified either (i) directly from the reaction mixtures using the Wizard^{tm} PCR Preps kit (Promega) or (ii) via preparative agarose gel electrophoresis with isolation of the dsDNA product employing the GeneClean^{tm} kit (Bio 101). The practical yields of these PCR products after 31 cycles with *Taq* DNA polymerase were similar regardless of the sense primer content.

DNA sequencing

These PCR products were sequenced via the dideoxy chain termination method in a cycle sequencing protocol employing sense strand primers (corresponding to positions -10 to 6 and -10 to 13 of the *hisF* gene) and the PCR-amplified antisense strand as a template. Direct dsDNA cycle sequencing of the antisense strand after PCR provided a determination of the percent incorporation of dAMP, dCMP, dGMP and dTMP opposite the universal base during the PCR reaction. The cycle sequencing was carried out with either (i) the fmol Sequencing System^{tm} (Promega) using a 5'-^{33}P-end-labeled primer and Wizard^{tm}-purified template, or (ii) the Thermosequenase^{tm} kit (Amersham) using [[alpha]-^{33}P] ddNTPs, unlabeled primer, and gel-purified template. The best signal/background ratios for this study were observed with the latter sequencing procedure. The manufacturers' protocols for the cycle sequencing reactions were employed with the exception of the annealing temperature (57°C) used in the thermal cycle.

Analysis of results

Quenched cycle sequencing reaction products were analyzed by gel electrophoresis under denaturing conditions (7 M urea, 50-55°C) on 8% polyacrylamide gels (0.4 mm * 30 cm * 40 cm) in TTE buffer (90 mM Tris-taurine, 0.5 mM EDTA) on a Gibco BRL model S2 apparatus at constant power (85 W). Gels poured using a molded silicone gel casting clamp (Gibco BRL) tended to result in significantly less 'smiling' of the gel during electrophoresis, which facilitated the quantitative analysis of the imaged data. The gels were preequilibrated for 1-2 h prior to loading samples. For best results, the samples (2.5 μl) were loaded either (i) in every second lane using sharkstooth combs, or (ii) in wells using a well comb. Gels were run until the bromophenol blue tracking dye had moved through 75% of the gel length. The gels were transferred from the glass sandwich to blotting paper (Whatman), covered with Saran Wrap^{tm}, and dried *in vacuo* on a Bio-Rad model 583 gel dryer for 2 h at 80°C.

The distributions of bases incorporated opposite the candidate universal bases in the PCR primer were quantified by phosphorimaging. Detection was carried out by 2-3 day exposures of dried sequencing gels on BI imaging plates (Bio-Rad). The image data was imported with a GS363 plate scanner (BioRad). Data work-up was performed on an Apple Macintosh 7500/100 using Molecular Analyst^{tm} software from Bio-Rad. A one-dimensional graphic profile extending horizontally across the sequencing gel at the position of azole base incorporation was extracted. Several background profiles in the vicinity of the experimental data were also extracted, averaged and subtracted from the experimental profile. The background-subtracted experimental profile was smoothed resulting in a trace with four peaks, corresponding to pixel density in the A, C, G and T lanes, which were subsequently integrated. The integration data were exported to Microsoft Excel for calculation of percent incorporation and error analysis. Note that the sense strand is being sequenced in this experiment, using antisense DNA as template. Pixel density in the A lane therefore corresponds to incorporation of dTMP opposite the template azole base by *Taq* DNA polymerase.

RESULTS

The distribution analysis of dNTP incorporation into replicating DNA strands is based upon DNA sequencing of primer-modified PCR products. Each azole base nucleoside was incorporated at position 36 in 40 base oligonucleotide PCR primers encoding positions -10 to 30 on the sense strand of *E.coli hisF* gene in the p*hisF-tac* vector (22) (Table 1). When used at low stringency annealing temperatures (37°C), all cases of the *hisF* PCR primers with modifications including inosine, abasic or azole nucleobases allowed for successful amplification of the template DNA by *Taq* DNA polymerase. Dideoxy cycle sequencing of the replicated antisense strand using a common sense strand primer allowed for quantitative assessment of the base composition via phosphorimaging of the resultant polyacrylamide gels. Figure 3 shows an example of the imaged DNA sequence data in which the direct comparison of a positive control (lane 1) can be made with those containing azole carboxamide nucleobases. The data in Table 2 represent averages from eight gel results from four independent sequencing reactions for each of nine distinct combinations of template and PCR primer. The validity of the analysis is supported by the even distribution of nucleotide incorporation observed when an equimolar mixture of four oligonucleotide primers, representing all four natural bases at position 36, are used as the sense strand primer. The mutagenic rate for base substitution errors by *Taq* DNA polymerase under the conditions of this assay are reported (23) to be in the range of $1/10^5$, and thus approximately three orders of magnitude less than the error (up to 9%) observed for this experimental approach.

Figure 3. Analysis of the base distribution via Sanger dideoxy sequencing of the antisense DNA for PCR products containing (**a**) a mixture of all four natural bases, 

(**b**) $\mathbf{PzA_3}$ and (**c**) $\mathbf{PzA_4}$ in the sense strand. The arrows mark the horizontal cross section of the gel at the point where the mixture of nucleotides was incorporated into the antisense DNA opposite the azole base. Note that in Table 2 the reported percentages of dNMPs incorporated opposite the azole bases were calculated as the inverse of the gel results (i.e., a band in the G lane on the sequencing gel was assumed to arise from the incorporation of dCMP opposite the azole base).

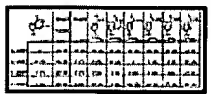


Table 2. Nucleotide incorporation directed by templates containing candidate The percentage incorporation of each dNMP by the *Taq* DNA polymerase. These values are the average of eight experiments, whereby the PCR, the dideoxy cycle sequencing reactions and polyacrylamide gel electrophoresis were performed in duplicate for each base-analog-containing oligonucleotide. The standard error of these values is given in parentheses.

The *Taq* DNA polymerase consistently incorporated dAMP at the specified position to the near exclusion of any other nucleotide in templates containing an abasic site. Under the same experimental conditions, the *Taq* DNA polymerase incorporated both dCMP and dAMP opposite hypoxanthine in approximately a 5.5:1 ratio. These results differ slightly from those of Kamiya *et al.* (4), where only incorporation of dCMP opposite hypoxanthine by *Taq* DNA polymerase was observed. However, the results reported here are consistent with the observed relative stabilities of I-C and I-A base pairs from thermodynamic measurements (24,25). As a contrast to the azole carboxamides, the base analogue $\mathbf{PrN_3}$ was designed to maximize stacking interactions, and not participate in specific hydrogen-bonding interactions with the naturally occurring bases (15). A relaxed degree of specificity was observed in this case since both dAMP and dTMP (in a 3:1 ratio, respectively) were incorporated opposite the nitropyrrole base $\mathbf{PrN_3}$ by *Taq* DNA polymerase.

The carboxamide side chain was used to substitute the pyrrole nucleus to give $\mathbf{PrA_3}$ on the premise that it may adopt multiple hydrogen bond donor-acceptor patterns through internal bond rotations (16). This modification resulted in a shift of the specificity toward T incorporation relative to the nitro substituted pyrrole $\mathbf{PrN_3}$. Addition of a second ring nitrogen into position 2 was not expected to present any additional hydrogen bond acceptor for base pairing but would affect the electronic properties relative to the azole base $\mathbf{PrA_3}$. This alteration in the heterocycle can be detected in the A and T ratio of incorporation by *Taq* DNA polymerase; template containing $\mathbf{PzA_4}$ favored A over T in contrast to $\mathbf{PrA_3}$. A simple position change for an endocyclic nitrogen alters the known electronegativity and dipole moment of the azole. This feature in $\mathbf{ImA_4}$ was predicted to impart at least two base pairing conformations that differ through intramolecular hydrogen bonds (16). However, the observed A to T incorporation ratios for the imidazole base $\mathbf{ImA_4}$ templates indicated that the directing properties of this base mirror those values observed for the pyrrole nucleus $\mathbf{PrA_3}$.

The results with pyrazole-3-carboxamide ($\mathbf{PzA_3}$) are a contrast to those discussed above. In this study, $\mathbf{PzA_3}$ behaved like a universal purine analogue, directing approximately the equivalent incorporation of dCMP and dTMP by *Taq* DNA polymerase (Table 2). In comparison to other purine analogs, dTMP was qualitatively observed (3,4) to be incorporated by the *Taq* DNA polymerase opposite 6-

methoxyguanine with greater frequency than dCMP. The *Taq* DNA polymerase also incorporated dTMP with much greater frequency than dCMP opposite 2-amino-6-methoxyaminopurine (11). As a template for *in vitro* DNA replication catalyzed by thermostable DNA polymerases, $\mathbf{PzA_3}$ appears to be the most promising candidate as a universal purine analogue examined to date.

DISCUSSION

The thermal stability of oligonucleotide sequences containing a single substitution of inosine has defined the role for this nucleoside as a probe for degenerate DNA hybridization (26,27). An expansion of this universal nucleoside recognition concept is represented by the use of PCR primers containing inosine and the various azole base analogs. The results using this assay system are interpreted for the application of the azole bases as degenerate templates in PCR. Hypoxanthine base appears to direct the incorporation of nucleotides in a manner consistent with its hydrogen bonding base-pairing attributes. However, the studies presented here indicate that the biochemical behavior of the azole nucleobase analogs in DNA templates are in contrast to inosine since they cannot be predicted solely on the basis of their ability to stabilize duplex DNA. All of the azole nucleobases bear a similar overall shape and size which allows for a comparison of the template effects within a single DNA sequence context. There are suprisingly dramatic effects on the template specificity revealed upon alteration of the heterocycle electronic properties. These simple nitrogen substitutions indicate a role for the electronic character of the azole nucleobase in dictating the selective incorporation of nucleotide triphosphates on the leading strand.

As a design approach, the azole nucleobases display important advantages for degenerate recognition in DNA replication. These analogs are fundamentally distinct from purine/pyrimidines whose base pairing is dictated by a precise balance of conformational and tautomeric equilibria. Subtle structural changes in the azole nucleobases result in large changes of the template recognition by *Taq* DNA polymerase consistent with a molecular event that is chemically tunable. The combined effects of ring electronics and a conformationally dynamic functional group expands the potential for nucleobase analog design. While none of the nucleobase analogs discussed here were fully degenerate, a prediction can be made on the basis of this study that useful incorporation of nucleotide mixtures at a single position by *Taq* DNA polymerase could be achieved and applied to oligonucleotide mediated site directed mutagenesis. It is fully anticipated that differences among DNA polymerases are potentially revealed by use of these nucleobase analogs as recently displayed for alternate hydrogen bonded base pairing patterns in another unique series (29). A full sequence context study is beyond the scope of this initial study, and there exists the possibility that the absolute base specificity could also be influenced by position. Regardless, the dependence of the azole base electronic character within a sequence context is significant impetus for further investigations of the physical, chemical and biological properties of DNA templates containing these nucleobase analogs.

One of the striking features of these results is the strong bias that *Taq* polymerase shows for inserting A opposite modified nucleobases or an abasic site. The tendency for A insertion opposite an abasic site or nucleobase analog by DNA polymerases has been observed previously, but the molecular basis for this effect remains poorly resolved (29-31). A theoretical hydrogen bonding scheme predicts that none of the modified bases included in this initial study, with the exception of inosine, should base pair selectively with A. In azole nucleobases, the simple nitropyrrole $\mathbf{PrN_3}$, and three of the carboxamide substituted derivatives, $\mathbf{PrA_3}$, $\mathbf{PzA_4}$ and $\mathbf{ImA_4}$ incorporate A at substantial levels. More important to the application of nucleobase analogs is the property displayed by $\mathbf{PzA_3}$ to overcome the strong preference for a DNA polymerase to insert A. These results provide a striking confirmation of our hypothesis that substituted carboxamide azoles are a unique class of compounds which show distinct, multiple

biological recognition features. The mechanisms by which these nucleobase analogs operate are not clearly established at this time. However, it is apparent that because the azole nucleobases all have similar shapes, the potential exists for establishing molecular parameters for DNA polymerase-template recognition through analysis of their structure-activity relationships.

Recently, the 2'-deoxyribonucleotide triphosphate derivative of $\mathbf{ImA_4}$ was investigated as a potential substrate for *Taq* DNA polymerase (32). Misincorporation frequencies were reported that are comparable to those of other hypermutagenic methods. The occurrence of tranversions as well as transitions suggest that alternate conformations of the imidazole nucleobase are important in enzyme recognition. In combination with the studies shown herein, these results establish a basis for the future development of azole-nucleobase analogs displaying useful molecular recognition properties in enzymatic methods for DNA synthesis.

ACKNOWLEDGEMENTS

We thank Thomas Klem for initial studies of the template properties of the base analogs as well as Doug Klewer for his insights during the development of the work presented herein. We also thank the NIH GM53155 for financial support.

REFERENCES

1 Koizume,S., Kamiya,H., Inoue,H. and Ohtsuka,E. (1994) *Nucleosides Nucleotides*, **13**, 1517-1534.

2 Kamiya,H., Ueda,T., Ohgi,T., Matsukage,A. and Kasai,H. (1995) *Nucleic Acids Res.*, **23**, 761-766. MEDLINE Abstract

3 Kamiya,H., Murata-Kamiya,N., Fujimuro,M., Kido,K., Inoue,H., Nishimura,S., Masutani,C., Hanaoka,F. and Ohtsuka,E. (1995) *Jpn J. Cancer Res.*, **86**, 270-276. MEDLINE Abstract

4 Kamiya,H., Sakaguchi,T., Murata,N., Fujimuro,M., Miura,H., Ishikawa,H., Shimizu,M., Inoue,H., Nishimura,S., Matsukage,A., Masutani,C., Hanaoka,F. and Ohtsuka,E. (1992) *Chem. Pharm. Bull.*, **40**, 2792-2795. MEDLINE Abstract

5 Singer,B., Chavez,F., Goodman,M.F., Essigmann,J.M. and Dosanjh,M.K. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8271-8274. MEDLINE Abstract

6 Eritja,R., Horowitz,D.M., Walker,P.A., Ziehler-Martin,J.P., Boosalis,M.S., Goodman,M.F., Itakura,K. and Kaplan,B.E. (1986) *Nucleic Acids Res.*, **14**, 8135-8153. MEDLINE Abstract

7 Grevatt,P.C., Solomon,J.J. and Bhanot,O.S. (1992) *Biochemistry*, **31**, 4181-4188. MEDLINE Abstract

8 Kong Thoo Lin,P. and Brown,D.M. (1989) *Nucleic Acids Res.*, **17**, 10373-10383.

9 Brown,D.M. and Kong Thoo Lin,P. (1991) *Carbohydrate Res.*, **216**, 129-139.

10 Nishio,H., Ono,A., Matsuda,A. and Ueda,T. (1992) *Nucleic Acids Res.*, **20**, 777-782. MEDLINE Abstract

11 Kamiya,H., Murata-Kamiya,N., Kong Thoo Lin,P., Brown,D.M. and Ohtsuka,E. (1994) *Nucleosides*

Nucleotides, **13**, 1483-1492.

12 Johnson,W.T., Zhang,P. and Bergstrom,D.E. (1997) *Nucleic Acids Res.*, **25**, 559-567. MEDLINE Abstract

13 Lawyer,F.C., Stoffel,S., Saiki,R.K., Myambo,K., Drummond,R. and Gelfand,D.H. (1989) *J. Biol. Chem.*, **264**, 6427-6437. MEDLINE Abstract

14 Tindall,K.R. and Kunkel,T.A. (1988) *Biochemistry*, **27**, 6008-6013. MEDLINE Abstract

15 Bergstrom,D.E., Zhang,P., Toma,P.H., Andrews,P.C. and Nichols,R. (1995) *J. Am. Chem. Soc.*, **117**, 1201-1209.

16 Bergstrom,D.E., Zhang,P. and Johnson,W.T. (1996) *Nucleosides*, **15**, 59-68.

17 Ramasamy,K., Robins,R.K. and Revankar,G.R. (1986) *Tetrahedron*, **42**, 5869-5878.

18 Srivastava,P.C., Robins,R.K., Takusagawa,F. and Berman,H.M. (1981) *J. Heterocycl. Chem.*, **18**, 1659-1662.

19 Rousseau,R.J., Robins,R.K. and Townsend,L.B. (1970) *J. Heterocycl. Chem.*, **7**, 367-372.

20 Chen.Z. and Ruffner,D.E. (1996) *BioTechniques*, **21**, 820-822. MEDLINE Abstract

21 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

22 Klem,T.J. and Davisson,V.J. (1993) *Biochemistry*, **32**, 5177-5186. MEDLINE Abstract

23 Eckert,K.A. and Kunkel,T.A. (1990) *Nucleic Acids Res.*, **18**, 3739-3744. MEDLINE Abstract

24 Martin,F.H., Castro,M.M., Aboul-ela,F. and Tinoco,I.,Jr (1985) *Nucleic Acids Res.*, **13**, 8927-8938. MEDLINE Abstract

25 Kawase,Y., Iwani,S., Inone,H., Miura,K. and Ohtsuka,E. (1986) *Nucleic Acids Res.*, **14**, 7727-7736. MEDLINE Abstract

26 Case-Green,S.C. and Southern,E.M. (1994) *Nucleic Acids Res.*, **22**, 131-136. MEDLINE Abstract

27 Ohtsuka,E., Matsuki,S., Ikehara,M., Takahashi,Y. and Matsubara,K. (1985) *J. Biol. Chem.*, **260**, 2605-2608. MEDLINE Abstract

28 Horlacher,J., Hottiger,M., Podust,V.N., Hübscher,U. and Benner,S.A. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6329-6333. MEDLINE Abstract

29 Strauss,B.S. (1991) *BioEssays*, **13**, 79-84. MEDLINE Abstract

30 Moran,S., Ren,R.X.-F., Rummey,S.,IV and Kool,E.T. (1997) *J. Am. Chem. Soc.*, **119**, 2056-2057.

31 Cai,H., Bloom,L.B., Eritja,R. and Goodman,M.F. (1993) *J. Biol. Chem.*, **268**, 23567-23572.